

CMB Likelihood Functions for Beginners and Experts

Andrew H. Jaffe¹, J.R. Bond², P.G. Ferreira^{3,4}, & L.E. Knox⁵

¹*Center for Particle Astrophysics, University of California, Berkeley, CA 94720,
CITA, Toronto, ON, Canada*

³*Theory Group, CERN, CH-1211, Geneve 23, Switzerland*

⁴*CENTRA, Instituto Superior Tecnico, Lisboa 1096 Codex, Portugal*

⁵*Astronomy & Astrophysics Center, U. Chicago, Chicago, IL*

Abstract. Although the broad outlines of the appropriate pipeline for cosmological likelihood analysis with CMB data has been known for several years, only recently have we had to contend with the full, large-scale, computationally challenging problem involving both highly-correlated noise and extremely large datasets ($N > 1000$). In this talk we concentrate on the beginning and end of this process. First, we discuss estimating the noise covariance from the data itself in a rigorous and unbiased way; this is essentially an iterated minimum-variance mapmaking approach. We also discuss the unbiased determination of cosmological parameters from estimates of the power spectrum or experimental bandpowers.

I INTRODUCTION

In principle, CMB anisotropy data analysis is easy: for at least some subset of theories (i.e., inflation) and some subset of CMB experiments, exact likelihood functions can be written down. In these cases, at least, the rest of the problem is implementation: the numerical calculation of the probability distribution of the interesting parameters given the data.

We model a measurement of the CMB as

$$d_i = \sum_p A_{ip} T_p + n_i \quad (1)$$

where d_i is the data at time t_i , T_p is the beam-smeared temperature at pixel p (pointed at by unit vector $\hat{\mathbf{x}}_p$), A_{ip} is the pointing operator such that $A_{ip} = 1$ when pixel p is observed at time i , and n_i is the noise at time i . The noise is characterised by its distribution, for our purposes assumed to be a Gaussian with covariance $\langle n_i n_{i'} \rangle = N_{ii'}$.

Given the data, then, we first wish to solve for the sky temperature T_p and its posterior distribution, and then use this map to determine the statistical properties of the CMB sky, such as its power spectrum, C_ℓ and its distribution, from which we will finally determine the underlying cosmological parameters.

Unfortunately, this implementation is indeed complicated in realistic cases. It is usually assumed that the error distribution for an experiment is somehow handed down to the theorists and data analysts by the experimenters, themselves with a direct link to some higher being who decides these things. Of course this is not actually the case: instead, the error distribution itself must be estimated from the data along with the signal. In some (not necessarily realistic) cases, this is simple. If it were known *a priori* that the noise were Gaussian and white, then $N_{ii'} = \sigma^2 \delta_{ii'}$ and we could use the histogram of observations at a given pixel to estimate the noise variance σ .

In more realistic cases, the noise in the timestream will be correlated and contaminated by systematic problems and glitches such as cosmic rays, making any simple procedure like this infeasible in practice. In this paper we outline a self-consistent Bayesian (likelihood) method for determining the noise power spectrum along with the CMB map.

Given a map and its statistical distribution (i.e., the error matrix for Gaussian noise), determining the power spectrum (for the case of a Gaussian CMB signal, as well) has been well-studied; problems remain in the implementation of efficient algorithms, but the ideal solution is known.

In this paper, we also address the question of what to do next: how do we go from an estimate of the CMB fluctuation power spectrum, C_ℓ , to the underlying cosmological parameters? Because of the complicated (non-Gaussian) nature of the likelihood function, we will show that simple χ^2 techniques will necessarily bias the determination of the parameters, but will derive an ansatz for eliminating this bias.

The plan of this paper is as follows. First, we discuss the CMB Data-Analysis Pipeline from the standpoint of likelihood functions. Then we address the specific problem of simultaneously determining the noise and signal from the timestream. Finally, we address the problem of estimating parameters from the power spectrum.

II LIKELIHOODS FOR CMB EXPERIMENTS

As has become customary, we start our analysis with Bayes' theorem

$$P(\theta|DI) \propto P(\theta|I)P(D|\theta I) \quad (2)$$

where θ are the parameters we are trying to determine, D is the data, and I is the “background information” describing the problem. The quantity $P(D|\theta I)$ is thus the likelihood, the probability of the data given a specific set of parameters, $P(\theta|I)$ is the prior probability for the parameters, and the left-hand side of the equation is the posterior probability for the parameters given the data.

As above and throughout, we will take the data, d_i , as given by a time series of CMB measurements,

$$d_i = s_i + n_i = \sum_p A_{ip} T_p + n_i, \quad (3)$$

where i labels the time, $t = i\delta t$, s_i and n_i are the experimental noise and sky signal contributions at that time. The signal is in turn given by the operation of a “pointing matrix,” A_{ip} on the sky signal at pixel p , T_p (i.e., the “map”); we take the latter to be already pixelized and smeared by the experimental beam, so A is a very sparse matrix with a single “1” entry for each time corresponding to the observed pixel. In the future we will often rely on the summation convention and write $\sum_p A_{ip} T_p = A_{ip} T_p$, or occasionally use matrix notation, as in AT , etc.

We will assume that the observed noise n_i is a realization of a stationary Gaussian process with power spectrum $\tilde{N}(\omega)$. This means that the correlation matrix of the noise is given by

$$\langle n_i n_{i'} \rangle = N_{ii'} = \int \frac{d\omega}{2\pi} \tilde{N}(\omega) e^{-i\omega\delta t(i-i')}. \quad (4)$$

The stationarity of the process requires (or is defined by) $N_{ii'} = N(t_i - t_i')$.

Most generally, we will take the parameters to be

- The observed CMB signal on the sky, T_p ;
- The power spectrum of the noise, $\tilde{N}(\omega)$
- (Possibly) any cosmological parameters which describe the distribution of the T_p (i.e., the CMB power spectrum, C_ℓ , although we could also use the cosmological parameters such as H_0 and Ω directly).

Sometimes, we will assign a prior distribution for the CMB signal such that the cosmological parameters will be irrelevant; other times we will marginalize over the CMB signal itself and determine those parameters.

With these parameters and the data, d_i , Bayes' theorem becomes

$$P[T_p, \tilde{N}, C_\ell | d_i, I] \propto P[\tilde{N}|I] P(T_p, C_\ell | I) P[d_i | \tilde{N}(\omega), T_p, I]. \quad (5)$$

Here, we have used two pieces of information to simplify slightly. First, the noise power spectrum, $N(\omega)$ does not depend at all on the signal, so we can separate out its prior distribution. Second, given the noise power spectrum and the sky signal, the likelihood does not depend upon the cosmological parameters. For the Gaussian noise we assume, the likelihood is simply

$$\begin{aligned} -2 \ln \mathcal{L} &\equiv -2 \ln P[d_i | \tilde{N}, T_p] \\ &= \ln |N_{ii'}| + \sum_{ii'} (d_i - s_i) N_{ii'}^{-1} (d_{i'} - s_{i'}) = \sum_k \left[\ln \tilde{N}_k + |\tilde{d}_k - \tilde{s}_k|^2 / \tilde{N}_k \right] \end{aligned} \quad (6)$$

(ignoring an additive constant); recall that $s_i = \sum_p A_{ip} T_p$. The second equality uses tildes to denote the discrete Fourier transform at angular frequency ω_k (see appendix and/or below...).

We will now apply these general formulae to various cases.

A Known noise power spectrum

We will start with the simplest case, where we have complete prior knowledge of the noise power spectrum. This is the case that has been previously discussed in the literature, but we emphasize that it is very unrealistic!

We assign a delta-function prior distribution to N , transforming it in effect from a parameter to part of the prior knowledge. First, we assume no cosmological information about the distribution of temperatures on the sky: $P(T_p, C_\ell | I) = P(T_p | I)P(C_\ell | I)$; with this separation the posterior for C_ℓ is simply the prior—the experiment gives us no new information. We will also assign a uniform prior to T_p , lacking further information. Now, the posterior distribution for the sky temperature is simply proportional to the likelihood, which can be rewritten by completing the square as

$$P(T_p | \tilde{N}, d, I) \propto P(d | T_p, \tilde{N}, I) = \frac{1}{|2\pi C_{Npp'}|^{1/2}} \exp \left[-\frac{1}{2} \sum_{pp'} (T_p - \bar{T}_p) C_{N,pp'}^{-1} (T_{p'} - \bar{T}_{p'}) \right] \quad (7)$$

with the mean (also, the likelihood maximum) given by

$$\bar{T} = (A^T N^{-1} A)^{-1} A^T N^{-1} d \quad (8)$$

(in matrix notation), and the noise correlation matrix by

$$C_N = (A^T N^{-1} A)^{-1}. \quad (9)$$

Occasionally, the inverse of this correlation matrix is referred to as the *weight matrix*. As is usual for linear Gaussian models, the mean is just the multidimensional least-squares solution to $d = AT$ with noise correlation N . This is just the mapmaking procedure advocated in Refs. [1,2], cast into the form of a Bayesian parameter-estimation problem.

For the case of known noise, however, this map is more than a just a visual representation of the data. Even if we wish to determine the cosmological parameters, it is an essential quantity. We can write the prior for both the map and the spectrum as

$$P(C_\ell, T_p | I) = P(T_p | C_\ell, I)P(C_\ell | I) \quad (10)$$

using the laws of probability, and so we can see that our rewriting of the likelihood in the form of Eq. 7 remains useful. That is, the full distribution is only a function of the data through the maximum-likelihood map, \bar{T} —in statistical parlance, \bar{T} is a *sufficient statistic*. Thus for known noise, we can *always* start by making a map (and calculating its noise matrix, C_N).

1 Cosmological CMB priors

Here, we briefly examine the specific form of the signal prior, $P(T_p | C_\ell, I)$, motivated by simple Gaussian models. That is, we take the sky temperature, T_p , to be an actual realization of a Gaussian CMB sky, with covariance specified by the power spectrum, C_ℓ ,

$$\langle T_p T_{p'} \rangle = C_{T,pp'} = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell B_\ell^2 P_\ell (\hat{x}_p \cdot \hat{x}_{p'}) \quad (11)$$

(note that we include beam-smearing by a symmetric beam with spherical harmonic transform B_ℓ in this definition); $\hat{x}_p \cdot \hat{x}_{p'}$ gives the cosine of the angle between the pixels. With this covariance, the prior becomes

$$P(T_p|C_\ell, I) = \frac{1}{|2\pi C_{Tpp'}|^{1/2}} \exp \left[-\frac{1}{2} \sum_{pp'} T_p C_{T,pp'}^{-1} T_{p'} \right]. \quad (12)$$

We thus have a posterior distribution for T_p and C_ℓ which is the product of two Gaussians. In the usual cosmological likelihood problem, we don't care about the actual sky temperature *per se*, but are concerned with the C_ℓ (or the parameters upon which the power spectrum depends). Thus, we can marginalize over the T_p ,

$$\begin{aligned} P(C_\ell|d, I) &= \int dT_p P(C_\ell, T_p|d, I) = P(C_\ell|I) \int dT_p P(T_p|C_\ell, I) P(d|T_p, I) = P(C_\ell|I) P(\bar{T}(d)|C_\ell I) \\ &= P(C_\ell|I) \frac{1}{|2\pi C_T|^{1/2}} \frac{1}{|2\pi C_N|^{1/2}} \exp \left[-\frac{1}{2} \sum_{pp'} \left(T_p C_{T,pp'}^{-1} T_{p'} + (T_p - \bar{T}_p) C_{N,pp'}^{-1} (T_{p'} - \bar{T}_{p'}) \right) \right] \end{aligned} \quad (13)$$

where we have included a prior for the power spectrum itself, so we can write the Gaussian factor as the likelihood for the map given C_ℓ , $P(C_\ell|d, I) \propto P(C_\ell|I) P(\bar{T}|C_\ell)$. The equation defines the effective likelihood for the map (\bar{T} , now considered as the data rather than the timestream itself, d), which is easily computed again by completing the square, giving

$$P(\bar{T}_p|C_\ell I) = \frac{1}{|2\pi (C_{Tpp'} + C_{Npp'})|^{1/2}} \exp \left[-\frac{1}{2} \sum_{pp'} \bar{T}_p (C_T + C_N)_{pp'}^{-1} \bar{T}_{p'} \right]. \quad (14)$$

This is just the usual CMB likelihood formula: the “observed map,” \bar{T}_p , is just the sum of two quantities (noise and signal) distributed as independent gaussians. Note again that the data only enters through the maximum likelihood map, \bar{T} , although that calculation is only implicit in this formula. Further, the power spectrum C_ℓ only enters through the signal correlation matrix, C_T , and in a very nonlinear way.

We can also play a slightly different game with the likelihood. If we retain the gaussian prior for the CMB temperature but *fix* the CMB power spectrum, we can estimate the map with this additional prior knowledge. We will again be able to complete the square in the exponential and see that T_p is distributed as a gaussian:

$$P(T_p|C_\ell, d, I) = \frac{1}{|2\pi C_W|^{1/2}} \exp \left[-\frac{1}{2} \chi^2(T_p|C_\ell, d, I) \right] \quad (15)$$

with

$$\chi^2(T_p|C_\ell, d, I) \equiv \sum_{pp'} (\bar{T}_p - (W\bar{T})_p) C_{W,pp'}^{-1} (\bar{T}_{p'} - (W\bar{T})_{p'}). \quad (16)$$

Now, the mean is given by

$$(W\bar{T}) = C_T (C_T + C_N)^{-1} \bar{T} \quad (17)$$

which is just the *Wiener Filter!* It has correlation matrix given by

$$C_W = C_T (C_T + C_N)^{-1} C_T. \quad (18)$$

Note that the maximum-likelihood map, \bar{T} still appears in these formulae, but it is no longer the maximum of the *posterior* distribution, now given by the Wiener filter, $W\bar{T}$.

This subsection has shown how many of the usual CMB data calculations can be seen as different uses of the Bayesian formalism:

- the least-squares map (seeing it as a “sufficient statistic”),
- the CMB cosmological-parameter likelihood function, and
- the Wiener filter map of the CMB signal.

The differences depend on what quantity is estimated (the map or the power spectrum) and what prior information is included.

III SIMULTANEOUS ESTIMATION OF THE NOISE AND SKY MAP

We now turn to the (realistic!) case where the noise covariance of the experiment are not known prior to receiving the data. That is, we must use the data itself to estimate the noise properties.

The previous subsection briefly outlined the Bayesian approach to CMB statistics, assuming a known noise power spectrum. Now, we will relax this assumption and approach the more realistic case when we must estimate both the experimental noise and the anisotropy of the CMB. We will take as our model a noise power spectrum of amplitude \bar{N}_α in bands numbered α , with a shape in each band given by a fixed function P_k with a width of n_α (the number of discrete modes in the band); we will usually take $P_k = \text{const}$ so \tilde{N} is piecewise constant. That is,

$$\tilde{N}(\omega_k) = \bar{N}_\alpha P_k. \quad (19)$$

We again assign a constant prior to the sky map, T_p . As the prior for the noise we will take $P(\bar{N}_\alpha) \propto 1/\bar{N}_\alpha^\nu$. With $\nu = 1$ and a single band, this is the usual Jeffreys prior advocated for “scale parameters” and the units on \bar{N}_α are irrelevant.

With this model and priors, the joint likelihood for the noise and the map becomes

$$\begin{aligned} P(T_p, \bar{N}_\alpha | d, I) &\propto \prod_\alpha \frac{1}{\bar{N}_\alpha^{\nu+n_\alpha/2}} \exp \left[-\frac{1}{2\bar{N}_\alpha} \sum_{k \in \alpha} \frac{1}{P_k} |\tilde{d}_k - \tilde{A}_{kp} T_p|^2 \right] \\ &= \prod_\alpha \frac{1}{\bar{N}_\alpha^{\nu+n_\alpha/2}} \exp \left[-\frac{1}{2\bar{N}_\alpha} \sum_{k \in \alpha} \frac{1}{P_k} |\tilde{\varepsilon}_k|^2 \right] \end{aligned} \quad (20)$$

where $k \in \alpha$ refers to a sum over modes in band α . We also define the estimate of the noise as $\varepsilon = d - AT$ for future use.

We can simultaneously solve for the maximum-probability noise and signal. Carrying out the derivatives gives

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial T_p} &= \sum_\alpha \left[\frac{1}{\bar{N}_\alpha} \sum_{k \in \alpha} \frac{1}{P_k} (d_k - A_{kp'} T_{p'}) A_{kp} \right] \\ &= (d - AT)^T N^{-1} A = \varepsilon^T N^{-1} A \end{aligned} \quad (21)$$

(switching between indices in Fourier space and matrix notation) and

$$\frac{\partial \ln \mathcal{L}}{\partial \bar{N}_\alpha} = -\frac{1}{2} \frac{1}{\bar{N}_\alpha} \left[(n_\alpha + 2\nu) - \frac{1}{\bar{N}_\alpha} \sum_{k \in \alpha} \frac{1}{P_k} |\tilde{d}_k - \tilde{A}_{kp'} T_{p'}|^2 \right] \quad (22)$$

Setting these to zero and solving, we then find we must simultaneously satisfy

$$T = (A^T N^{-1} A)^{-1} A^T N^{-1} d \quad (23)$$

(using matrix notation for simplicity) and

$$\bar{N}_\alpha = \frac{1}{n_\alpha + 2\nu} \sum_{k \in \alpha} \frac{1}{P_k} |\tilde{d}_k - \tilde{A}_{kp'} T_{p'}|^2 = \frac{1}{n_\alpha + 2\nu} \sum_{k \in \alpha} \frac{1}{P_k} |\tilde{\varepsilon}_k|^2. \quad (24)$$

Equation 23 is just the usual maximum-likelihood map solution; Equation 24 is the average “periodogram” of the noise over the band α , with a slight modification for the prior probability, parameterized by ν ; for wide bands this is irrelevant. As we will see below, iteration is actually a very efficient way to solve these simultaneously. For future reference, we also write down the derivative with respect to the map at the joint maximum (i.e., substituting Eq. 24 into Eq. 21),

$$\frac{\partial \ln \mathcal{L}}{\partial T_p} = \sum_\alpha \left[(n_\alpha + 2\nu) \frac{\sum_{k \in \alpha} \tilde{\varepsilon}_k \tilde{A}_{kp}/P_k}{\sum_{k' \in \alpha} |\tilde{\varepsilon}_{k'}|^2 / P_{k'}} \right] \quad (25)$$

If we fix the noise at the joint maximum, then the problem reduces to that of the previous section, a Gaussian likelihood in T_p with which the usual tools can be applied. This is not a rigorous approach to the problem, however due to correlation between noise and signal estimation. To get a handle on this, we calculate the curvature of the distribution around this joint maximum, which we define as

$$\mathcal{F} = \begin{pmatrix} \mathcal{G}_{pp'} & \mathcal{G}_{kp} \\ \mathcal{G}_{pk} & \mathcal{G}_{kk'} \end{pmatrix} \quad (26)$$

where a subscript α or p refers to a derivative with respect to \bar{N}_α or T_p , respectively. \mathcal{F} refers to the full matrix; \mathcal{G} to the sub-blocks. Explicitly, the parameter derivatives are

$$\frac{\partial^2 \ln \mathcal{L}}{\partial T_p \partial T_{p'}} = \mathcal{G}_{pp'} = - \sum_{\alpha} \frac{1}{\bar{N}_\alpha} \sum_{k \in \alpha} \frac{\tilde{A}_{kp} \tilde{A}_{kp'}}{P_k} = -A^T N^{-1} A \quad (27)$$

and

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \bar{N}_\alpha \partial \bar{N}_{\alpha'}} = \mathcal{G}_{\alpha\alpha'} = -\frac{n_\alpha + 2\nu}{2\bar{N}_\alpha^2} \delta_{\alpha\alpha'} \quad (28)$$

with the cross-curvature

$$\frac{\partial^2 \ln \mathcal{L}}{\partial T_p \partial \bar{N}_\alpha} = \mathcal{G}_{p\alpha} = -\frac{1}{\bar{N}_\alpha^2} \sum_{k \in \alpha} \frac{\tilde{\varepsilon}_k \tilde{A}_{kp}}{P_k} \quad (29)$$

The distribution about the joint maximum is *not* a Gaussian in the \bar{N}_α directions, so there is more information available than this. Nonetheless, if we treat the distribution as if it were Gaussian, we can calculate the covariance matrix, given by the inverse of this curvature matrix:

$$\mathcal{F}^{-1} = \begin{bmatrix} (\mathcal{G}_{pp'} - \mathcal{G}_{p\alpha} \mathcal{G}_{\alpha\alpha'}^{-1} \mathcal{G}_{\alpha'p'})^{-1} & (\mathcal{G}_{op} - \mathcal{G}_{\alpha\alpha'} \mathcal{G}_{\alpha'p'}^{-1} \mathcal{G}_{p'p})^{-1} \\ (\mathcal{G}_{p\alpha} - \mathcal{G}_{pp'} \mathcal{G}_{p'\alpha'}^{-1} \mathcal{G}_{\alpha'\alpha})^{-1} & (\mathcal{G}_{\alpha\alpha'} - \mathcal{G}_{op} \mathcal{G}_{pp'}^{-1} \mathcal{G}_{p'\alpha})^{-1} \end{bmatrix} \quad (30)$$

In particular, the effective variance of the map is increased from C_N to

$$\begin{aligned} C_{Npp'}^{\text{eff}} &= (C_N^{-1} - \mathcal{G}_{p\alpha} \mathcal{G}_{\alpha\alpha'}^{-1} \mathcal{G}_{\alpha'p'})^{-1} \\ &= \left[\sum_{\alpha} \frac{1}{\bar{N}_\alpha} \sum_{k \in \alpha} \frac{\tilde{A}_{kp} \tilde{A}_{kp'}}{P_k} - 2 \sum_{\alpha} \frac{1}{(n_\alpha + 2\nu) \bar{N}_\alpha^2} \sum_{k \in \alpha} \frac{\tilde{\varepsilon}_k \tilde{A}_{kp}}{P_k} \sum_{k' \in \alpha} \frac{\tilde{\varepsilon}_{k'} \tilde{A}_{k'p'}}{P_{k'}} \right]^{-1} \end{aligned} \quad (31)$$

which we have evaluated at the simultaneous peak of the distribution to eliminate \bar{N}_α .

In Ref. [3], we examine this correction to C_N^{-1} (the weight matrix) more closely; for sufficiently wide bands, relative to the number of pixels in the map, the correction is negligible.

A Noise Marginalization

Instead of this joint solution, formally, at least, we know the appropriate procedure: marginalize over the quantity we don't care about (the noise power spectrum) to obtain the distribution for the quantity we wish to know (the map, T_p). We can actually carry out this integral in this case:

$$P(T_p | d_i, I) = \int d\tilde{N}_k P(T_p, N_k | d_i, I) \propto \prod_{\alpha} \left[\sum_{k \in \alpha} \frac{1}{P_k} \left| \tilde{d}_k - \tilde{A}_{kp} T_p \right|^2 \right]^{-(n_\alpha/2+\nu-1)} \quad (32)$$

(this is just Student's t distribution in a slightly different form than is usually seen). If we stay in Fourier space, the maximum of this distribution can be calculated to be the solution of

$$\frac{\partial \ln P}{\partial T_p} = \sum_{\alpha} \left[(n_{\alpha} + 2\nu - 2) \frac{\sum_{k \in \alpha} \frac{1}{P_k} (d_k - A_{kp'} T_{p'}) \tilde{A}_{kp}}{\sum_{k' \in \alpha} \frac{1}{P_{k'}} |d_{k'} - \tilde{A}_{k'p'} T_{p'}|^2} \right] \quad (33)$$

Note that this is *exactly the same form* as Eq. 25, the equation for the maximum probability map in the joint estimation case, with the prefactor $(n_{\alpha} + 2\nu)$ replaced by $(n_{\alpha} + 2\nu - 2)$. This is equivalent to changing the exponent of the prior probability from ν to $\nu - 1$: the marginalized maximum for $\nu = 1$ (Jeffery's prior) is the same as the joint maximum for $\nu = 0$ (constant prior). We have also seen that for $n_{\alpha} \gg 1$, the value of ν is irrelevant, so that these should be nearly equal. (Moreover, the numerical tools to calculate the joint solution, outlined below, can be used in this case, too.) Note also that if $n_{\alpha} + 2\nu - 2 \leq 0$, the equation isn't solved for any (finite) map. In this case, either our prior information is so unrestrictive or the bands are so narrow that it is impossible to distinguish between noise and signal, and the probability distribution has no maximum when the noise is marginalized.

In principle, any further analysis of the map would have to rely on the full distribution of Eq. 32. In practice, the t-distribution is quite close to a Gaussian [again, for $n_{\alpha} \gg 1$ —is this right?] (although the tails are suppressed by a power-law rather than an exponential). It will thus be an excellent approximation to take the distribution to be a Gaussian

$$P(T_p | d_i, I) \approx \frac{1}{2\pi C_{N,pp'}^{\text{eff}}}^{1/2} \exp \left[-\frac{1}{2} \sum_{pp'} (T_p - \bar{T}_p) (C_{N,pp'}^{\text{eff}})^{-1} (T_{p'} - \bar{T}_{p'}) \right] \quad (34)$$

with noise covariance given by

$$(C_{N,pp'}^{\text{eff}})^{-1} = -\frac{\partial^2 \ln P(T_p | d_i, I)}{\partial T_p \partial T_{p'}} = \sum_{\alpha} \left[\frac{(n_{\alpha} + 2\nu - 2)}{\sum_{k \in \alpha} |\tilde{\varepsilon}_k|^2} \sum_{k \in \alpha} \frac{\tilde{A}_{kp} \tilde{A}_{kp'}}{P_k} - 2 \frac{(n_{\alpha} + 2\nu - 2)}{(\sum_{k \in \alpha} |\tilde{\varepsilon}_k|^2)^2} \sum_{k \in \alpha} \frac{\tilde{\varepsilon}_k \tilde{A}_{kp}}{P_k} \sum_{k' \in \alpha} \frac{\tilde{\varepsilon}_{k'} \tilde{A}_{k'p'}}{P_{k'}} \right] \quad (35)$$

where \bar{T} is the solution to Eq. 33 and the derivatives are taken with respect to the *full* distribution of Eq. 33). Note that the inverse of C_N^{eff} (the weight matrix) is given again by the sum of two terms. The first is just the weight matrix that would be assigned given the maximum probability solution for the noise, Eq. 9 (with the numerically irrelevant change of the prior $\nu \rightarrow \nu - 1$ as noted above). Again we find a correction term due to the fact that we are only able to *estimate* this noise power spectrum, where the correction has exactly the same form as in the non-marginalized case (Eq. 31) above with a change of the prefactor.

In Ref. [3], we show how the simultaneous solution for noise and map can be performed iteratively. As we have shown, this is also the solution for the case when the noise is marginalized over.

B Discussion

Happily, the bottom line of this analysis is that the mapmaking phase of CMB data analysis can proceed as has been discussed by other authors in the past. We do recommend two slight additions to the pipeline:

- Iterative map & noise determination
 1. Start by assuming the data stream is all noise.
 2. Calculate the noise power spectrum via FFT.
 3. Inverse FFT, and use this noise matrix for mapmaking
 4. Subtract the estimated signal contribution from the map, and repeat.
- Calculate the correction to the noise matrix (Eq. 35) (and either use the correction or make sure it is negligible).

We note that it is only necessary to solve for the map itself during the iterative process, but not to invert the inverse noise matrix, C_N^{-1} . The former can, in principle, be accomplished in far fewer operations than a matrix inverse, so the iterative procedure need not add a great amount of computing time to the process.

However, it is conventional wisdom that an ideal observational strategy has signal to noise of unity. In this regime, the signal to noise *in the timestream* is very low, and a good approximation to the estimator of the noise power spectrum is the banded periodogram of the data stream itself. Thus one is able to avoid the time-consuming process of performing multiple iterations.

IV RADICAL COMPRESSION: FROM THE POWER SPECTRUM TO COSMOLOGICAL PARAMETERS

Above, we discussed how to make a map of the CMB and estimate its noise properties. Elsewhere, the estimation of the CMB anisotropy power spectrum (C_ℓ) from a map and its noise estimate have been described and implemented at least for data sets of moderate size. But the ultimate goal, of course, is to go from the power spectrum to the cosmological parameters. Various groups [4] have calculated how well future experiments are expected to perform; here, we describe what is necessary to begin to achieve such results.

The basic problem is that a plot power spectrum estimates and associated errors do not suffice to describe the likelihood function of the experiment. Usually, we plot $\mathcal{C} \equiv \ell(\ell+1)C_\ell/(2\pi) \pm \sigma_\ell$ and are tempted to do a χ^2 procedure using those errors and possibly covariances between ℓ to fit theoretical curves to the estimates. As we have seen above, however, the shape of the likelihood as a function of the C_ℓ is not Gaussian, implicitly assumed in the use of a χ^2 procedure.

One solution is simply to use a full likelihood-minimization procedure on the cosmological parameters given the entire dataset. That is, evaluate Eq. 14 directly for points throughout the cosmological parameter space (over which C_ℓ and thereby the correlation matrix C_T will vary). Since all calculations of the likelihood function to date require $O(N_{\text{pix}}^3)$ operations in general, this is prohibitive for experiments of even moderate size (especially if the likelihood must be searched over something like the 11-dimensional space appropriate for inflation-inspired models).

Instead, we must find efficient ways to approximate the likelihood function based on minimal information. In the rest of this paper, we discuss approximations motivated by our knowledge of the likelihood function. Each requires only knowledge of the likelihood peak and curvature (or variance) as well as a third quantity related to the noise properties of the experiment. Alternately, for already-calculated likelihood functions, each approximation gives a functional form for fitting with a small number of parameters.

A The solution: approximating the likelihood

1 Offset lognormal distribution

We already know enough about the likelihood to see a solution to this problem. For a given multipole ℓ , there are two distinct regimes of likelihood. Consider a simple all-sky experiment with uniform pixel noise and some symmetric beam. Taking such a map as our data, we write $\bar{T}_p = T_p + n_p$ which we transform to spherical harmonics, so the data has contributions from the signal, $a_{\ell m}$, and the noise, $n_{\ell m}$, and the likelihood is

$$-2 \ln P(\bar{T} | \mathcal{C}_\ell) = \sum_\ell (2\ell+1) \left[\ln (\mathcal{C}_\ell B_\ell^2 + \mathcal{N}_\ell) + \frac{\widehat{\mathcal{D}}_\ell}{\mathcal{C}_\ell B_\ell^2 + \mathcal{N}_\ell} \right] \quad (36)$$

(up to an irrelevant additive constant; cf. Eq. 14), with $\mathcal{N}_\ell \equiv \ell(\ell+1)N_\ell/(2\pi)$, where $N_\ell = \langle |n_{\ell m}|^2 \rangle$ is the noise power spectrum in spherical harmonics, and $\widehat{\mathcal{D}}_\ell \equiv [\ell(\ell+1)/(2\pi)] \sum_m |a_{\ell m}|^2 / (2\ell+1)$ is the power spectrum of the full data (noise plus beam-smoothed signal); we have written it as a different symbol from above to emphasize the inclusion of noise and again use script lettering to refer to quantities multiplied by $\ell(\ell+1)/(2\pi)$.

Now, the likelihood is maximized at $\mathcal{C}_\ell = (\widehat{\mathcal{D}}_\ell - \mathcal{N}_\ell)/B_\ell^2$ and the curvature about this maximum is given by

$$\mathcal{F}_{\ell\ell'}^{(C)} = -\frac{\partial^2 \ln P(\bar{T} | \mathcal{C}_\ell)}{\partial \mathcal{C}_\ell \partial \mathcal{C}_{\ell'}} = \frac{2\ell+1}{2} \left(\mathcal{C}_\ell + \mathcal{N}_\ell/B_\ell^2 \right)^{-2} \delta_{\ell\ell'} \quad (37)$$

so the error (defined by the variance) on a \mathcal{C}_ℓ is

$$\delta \mathcal{C}_\ell = \left(\mathcal{C}_\ell + \mathcal{N}_\ell/B_\ell^2 \right) / \sqrt{\ell+1/2}. \quad (38)$$

Note that in this expression there is once again indication of a bias if we assume Gaussianity: upward fluctuations have larger uncertainty than downward fluctuations. But this is not true for Z_ℓ where Z_ℓ is defined so that $\delta Z_\ell \propto \delta \mathcal{C}_\ell / (\mathcal{C}_\ell + \mathcal{N}_\ell / B_\ell^2)$. More precisely, $Z_\ell \equiv \ln(\mathcal{C}_\ell + \mathcal{N}_\ell / B_\ell^2)$. Since δZ_ℓ is proportional to a constant, our approximation to the likelihood is to take Z_ℓ as normally distributed. That is, we approximate

$$-2 \ln P(\bar{T} | \mathcal{C}_\ell) = \sum_{\ell\ell'} Z_\ell \left(M^{(Z)} \right)_{\ell\ell'}^{-1} Z_{\ell'} \quad (39)$$

(up to a constant) where $M_{\ell\ell'}^{(Z)} = (\mathcal{C}_\ell + x_\ell) M_{\ell\ell'}^{(C)} (\mathcal{C}_{\ell'} + x_{\ell'})$ where $M^{(C)}$ is the covariance matrix of the \mathcal{C}_ℓ , usually taken to be the inverse of the curvature matrix. We refer to Eq. 39 as the offset lognormal distribution of \mathcal{C}_ℓ . Somewhat more generally we write

$$Z_\ell = \ln(\mathcal{C}_\ell + x_\ell) \quad (40)$$

for some constant x_ℓ , which for the case in hand is $x_\ell = \mathcal{N}_\ell / B_\ell^2$.

It is illustrative to derive the quantity Z_ℓ in a somewhat more abstract fashion. We wish to find a change of variables from \mathcal{C}_ℓ to Z_ℓ such that the curvature matrix is a constant:

$$\frac{\partial \mathcal{F}_{\ell\ell'}^{(Z)}}{\partial Z_\ell} = 0. \quad (41)$$

That is, we want to find a change of variables such that

$$\left(\mathcal{F}^{(Z)} \right)_{LL'}^{-1} = \sum_{\ell\ell'} \frac{\partial Z_L}{\partial \mathcal{C}_\ell} \left(\mathcal{F}^{(C)} \right)_{\ell\ell'}^{-1} \frac{\partial Z_{L'}}{\partial \mathcal{C}_{\ell'}} \quad (42)$$

is not a function of Z . We immediately know of one such transformation which would seem to do the trick:

$$\frac{\partial Z_L}{\partial \mathcal{C}_\ell} = \mathcal{F}_{L\ell}^{1/2} \quad (43)$$

where the $1/2$ indicates a Cholesky decomposition or Hermitian square root. In general, this will be a horrendously overdetermined set of equations, N^2 equations in N unknowns. However, we can solve this equation in general if we take the curvature matrix to be given everywhere by the diagonal form for the simplified experiment we have been discussing (Eq. 37). In this case, the equations decouple and lose their dependence on the data, becoming (up to a constant factor)

$$\frac{dZ_\ell}{d\mathcal{C}_\ell} = (\mathcal{C}_\ell + \mathcal{N}_\ell / B_\ell^2)^{-1}. \quad (44)$$

The solution to this differential equation is just what we expected,

$$Z_\ell = \ln(\mathcal{C}_\ell + \mathcal{N}_\ell / B_\ell^2) \quad (45)$$

with correlation matrix

$$\left(\mathcal{F}^{(Z)} \right)_{\ell\ell'}^{-1} = \frac{\left(\mathcal{F}^{(C)} \right)_{LL'}^{-1}}{(\mathcal{C}_L + \mathcal{N}_L / B_L^2)(\mathcal{C}_{L'} + \mathcal{N}_{L'} / B_{L'}^2)} \quad (46)$$

where $\ell = L$ and $\ell' = L'$. Please note that we are calculating a constant correlation matrix; the \mathcal{C}_ℓ in the denominator of this expression should be taken at the peak of the likelihood (i.e., the estimated quantities).

We emphasize that, even for an all-sky continuously and uniformly sampled experiment (for which Eq. 37 is exact), this Gaussian form, Eq. 39, is only an approximation, since the curvature matrix is given by Eq. 37 only at the peak. Nonetheless we expect it to be a better approximation than a naive Gaussian in \mathcal{C}_ℓ (which we note is the limit $x_\ell \rightarrow \infty$ of the offset lognormal).

We have also considered another ansatz known as the equal variance approximation which we have found sometimes reproduces the shape of the likelihood better than the offset lognormal form above, especially in the case of upper limits on \mathcal{C}_ℓ .

B Noise power

In the above, we have derived the extra quantity needed in this ansatz, $x_\ell = \mathcal{N}_\ell$, but this was explicitly done only for the unrealistic case of an experiment looking at the whole sky with uniform noise, with a likelihood function Eq. 36. Can we generalize this to more complicated experiments, which have the more complicated likelihood function Eq. 14, with correlation matrices encoding such effects as

- nonuniform noise;
- correlated noise;
- incomplete sky coverage; and/or
- complicated chopping schemes?

Happily, the answer to this question is that we can, and have successfully applied the results to experiments as diverse as COBE/DMR and Saskatoon.

For a map with noise correlation matrix C_N , and signal correlation matrix C_T , we generalize our definition of x to

$$\mathcal{C}_B/x_B = \sqrt{\frac{\text{Tr}(C_N^{-1}C_{T,B}C_N^{-1}C_{T,B})}{\text{Tr}(C_T^{-1}C_{T,B}C_T^{-1}C_{T,B})}} \simeq \sqrt{\frac{\text{Tr}(C_N^{-1}C_{T,B}C_N^{-1}C_{T,B})}{\text{Tr}(C^{-1}C_{T,B}C^{-1}C_{T,B})}} - 1 \quad (47)$$

where we use B to label bands of ℓ rather than individual multipoles, so that \mathcal{C}_ℓ has a fixed shape in each band with amplitude \mathcal{C}_B and $C_{T,B} = \partial C_T / \partial \mathcal{C}_B$ is the derivative of the signal correlation matrix with respect to the power in the band. The second expression would be appropriate for a full-sky experiment where the signal and noise separate in the traces; it may be numerically easier to calculate due to zero or negative eigenvalues in C_T that often appear due to approximation. Other expressions that may be more appropriate for particular experimental strategies are given in Ref. [5].

C Tests and Results

Because exact likelihoods have been calculated for many different experiments to date, we can check this ansatz in great detail. (In the future, this will probably be impossible because the exact calculation of the likelihood is so expensive that other techniques for finding the maximum will be necessary [6].)

1 COBE/DMR

First, we compare with the likelihood for COBE/DMR at several values of ℓ . In Fig. 1 we show the actual likelihood for the COBE quadrupole and other multipoles, along with the Gaussian that would be assumed given the curvature matrix calculated from the data. The figures show another way of understanding the bias introduced by assuming Gaussianity: upward deviations from the mean (which is not actually the mean of the non-Gaussian distribution, but the mode) are overly disfavored by the Gaussian distributions while downward ones are overly probable. For example, the standard-CDM value of $\mathcal{C}_2 = 770\mu\text{K}^2$ is only 0.2 times less likely than the most likely value of $150\mu\text{K}^2$ but it seems like a 5-sigma excursion (4×10^{-6} times less likely) based on the curvature alone.

Although it is extremely pronounced in the case of the quadrupole this is a problem that plagues all CMB data: the actual distribution is skewed to allow larger positive excursions than negative. The full likelihood “knows” about this and in fact takes it into account; however, if we compress the data to observed $\mathcal{C}_\ell \pm \sigma_\ell$ (or even observed \mathcal{C}_ℓ and a correlation matrix $M_{\ell\ell'}$) we lose this information about the shape of the likelihood function. Because of its relation to the well-known phenomenon of cosmic variance, we choose to call this problem one of *cosmic bias*.

We emphasize that cosmic bias can be important even in high-S/N experiments with many pixels. We might expect the central limit theorem to hold in this case and the distributions to become Gaussian. Indeed they do, at least near the peak. However, the central limit theorem does not guarantee that the tails of the distribution

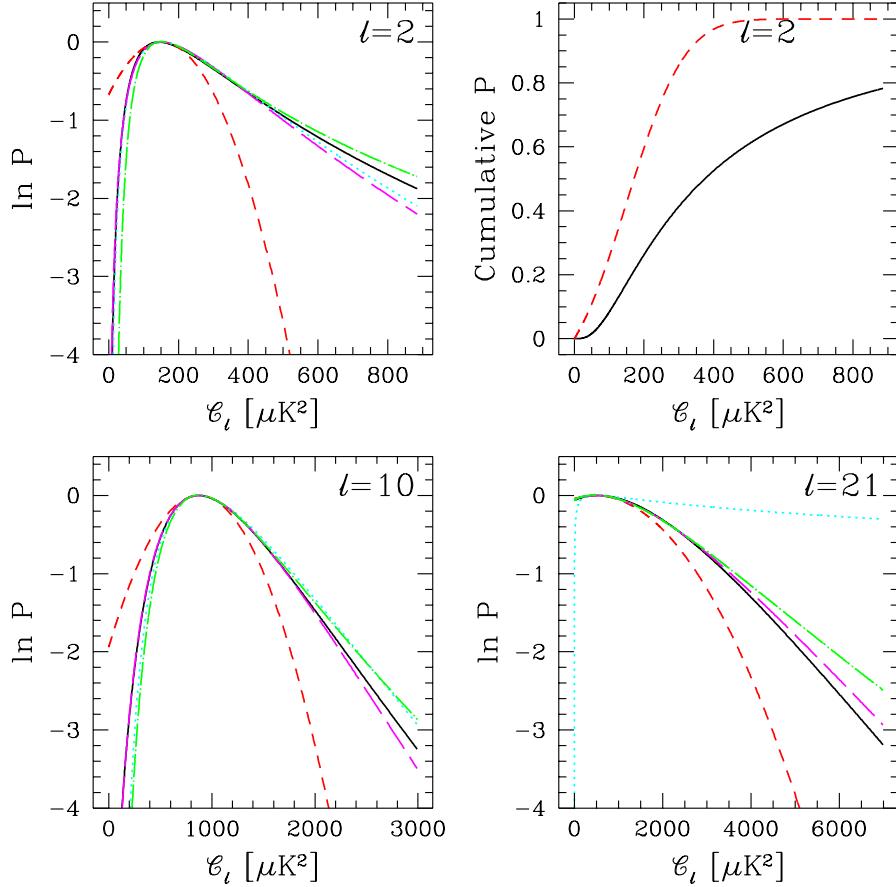


FIGURE 1. Full and approximate COBE/DMR likelihoods $P(\Delta|C_\ell)$ for various values of ℓ , as marked. The horizontal axis is $C_\ell = \ell(\ell+1)C_\ell/(2\pi)$. The upper right panel gives the cumulative probability. The solid (black) line is the full likelihood calculated exactly. The short-dashed (red) line is the Gaussian approximation about the peak. The dotted (cyan) line is a Gaussian in $\ln C_\ell$; the dashed (magenta) line is a Gaussian in $\ln(C_\ell + x_\ell)$, as discussed in the text. The dot-dashed (green) line is the equal-variance approximation.

will be Gaussian and there is the danger that a few seemingly discrepant points are given considerably more weight than they deserve. Cosmic bias has also been noted in previous work [7–9].

Putting the problem a bit more formally, we see that even in the limit of infinite signal-to-noise we cannot use a simple χ^2 test on \mathcal{C}_ℓ estimates; such a test implicitly assumes a Gaussian likelihood. Unlike the distribution discussed here, a Gaussian would have constant curvature ($\delta\mathcal{C}_\ell = \text{constant}$), rather than $\delta\mathcal{C}_\ell \propto \mathcal{C}_\ell$ as illustrated here.

2 Saskatoon

We apply this ansatz to the Saskatoon experiment, perhaps the apotheosis of a complicated chopping experiment. The Saskatoon data are reported as complicated chopping patterns (*i.e.*, beam patterns, H , above) in a disk of radius about 8° around the North Celestial Pole. The data were taken over 1993–1995 (although we only use the 1994–1995 data) at an angular resolution of $1.0\text{--}0.5^\circ$ FWHM at approximately 30 GHz and 40 GHz. More details can be found in Refs. [10,11]. The combination of the beam size, chopping pattern, and sky coverage mean that Saskatoon is sensitive to the power spectrum over the range $\ell = 50\text{--}350$. The Saskatoon dataset is calibrated by observations of supernova remnant, Cassiopeia–A. Leitch and collaborators [12] have recently measured the flux and find that the remnant is 5% brighter than the previous best determination. We have renormalized the Saskatoon data accordingly.

We calculated \mathcal{C}_ℓ for this dataset in [7]. We combine these results with the data’s noise matrix to calculate the appropriate correlation matrixes (in this case, the full curvature matrix) for Saskatoon and hence the appropriate x_B (Eq. 47) and thus our approximations to the full likelihood. In Figure 2, we show the full likelihood, the naive Gaussian approximation, and our present offset lognormal and equal-variance forms. Again, both approximations reproduce the features of the likelihood function reasonably well, even into the tails of the distribution, certainly better than the Gaussian approximation. They seem to do considerably better in the higher- ℓ bands; even in the lower ℓ bands, however, the approximations result in a *wider* distribution which is preferable to the narrower Gaussian and its resultant strong bias. Moreover, we have found that we are able to reproduce the shape of the true likelihood essentially perfectly down to better than “three sigma” if we simply *fit* for the x_B (but of course this can only be done when we have already calculated the full likelihood—precisely what we are trying to avoid!). For existing likelihood calculations, this method can provide better results without any new calculations (see [5] for our recommendations for the reporting of CMB bandpower results for extant, ongoing, and future experiments).

We have found that the shape of the power spectrum used with each bin of ℓ can have an impact on the likelihood function evaluated using this ansatz. Similarly, a finer binning in ℓ will reproduce the full likelihood more accurately. Although the maximum-likelihood amplitude at a fixed shape (n_s) does not significantly depend on binning or shape, the shape of the likelihood function along the maximum-likelihood ridge changes with finer binning and with the assumed spectral shape.

3 COBE/DMR + Saskatoon

As a further example and test of these methods, we can combine the results from Saskatoon and COBE/DMR in order to determine cosmological parameters. For this example, we use the orthogonal linear combinations as described in the previous section. In Figure 3 we show the likelihood contours for standard CDM, varying the scalar slope n_s and amplitude σ_8 . As before, we see that the naive χ^2 procedure is biased toward low amplitudes at fixed shape (n_s), but that our new approximation recovers the peak quite well. The full likelihood gives a global maximum at $(n_s, \sigma_8) = (1.15, 1.67)$, and our approximation at $(1.13, 1.58)$, while the naive χ^2 finds it at $(1.21, 1.55)$, outside even the three-sigma contours for the full likelihood. We can also marginalize over either parameter, in which case the full likelihood gives $n_s = 1.17_{-0.07}^{+0.08}$, $\sigma_8 = 1.68_{-0.21}^{+0.26}$; our ansatz gives $n_s = 1.14_{-0.05}^{+0.07}$, $\sigma_8 = 1.60 \pm 0.15$; and the naive χ^2 gives $n_s = 1.21_{-0.09}^{+0.08}$, $\sigma_8 = 1.55_{-0.20}^{+0.18}$. (Note that even with the naive χ^2 we marginalize by explicit integration, since the shape of the likelihood in parameter space is non-Gaussian in all cases.)

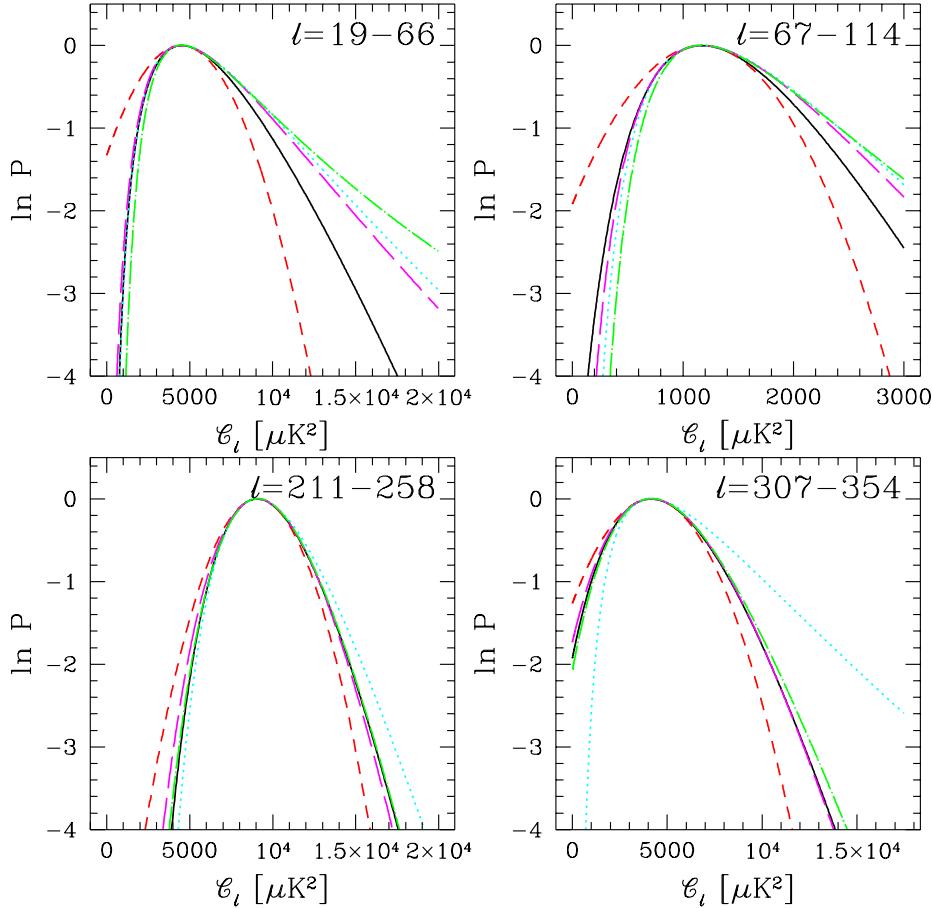


FIGURE 2. Full and approximate Saskatoon likelihoods. As in Fig. 1. The solid (black) line is the full likelihood calculated exactly. The short-dashed (red) line is the Gaussian approximation about the peak. The dotted (cyan) line is a Gaussian in $\ln C_\ell$; the dashed (magenta) line is a Gaussian in $\ln(C_\ell + x_\ell)$, as discussed in the text. The dot-dashed (green) line is the equal-variance approximation.

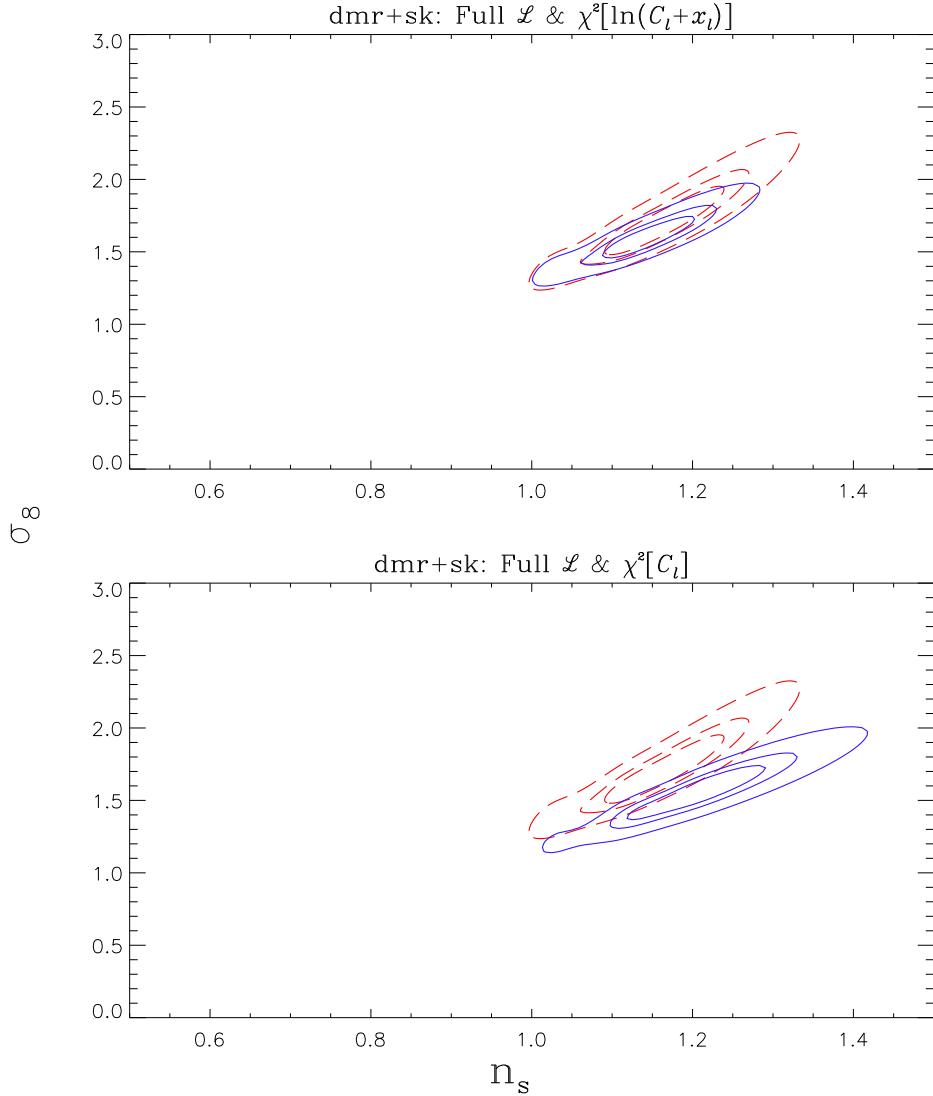


FIGURE 3. Likelihood contours for COBE/DMR and Saskatoon combined for the cosmological parameters n_s and σ_8 (with otherwise standard CDM values) combining likelihoods from COBE/DMR and Saskatoon. Contours are for ratios of the likelihood to its maximum equal to $\exp -\nu^2/2$ with $\nu = 1, 2, 3$. Upper panel is for the full likelihood (dashed) and its offset lognormal approximation as a Gaussian in $\ln(C_\ell + x_\ell)$ (solid; see text); lower panel shows the full likelihood and its approximation as a Gaussian in C_ℓ .

4 Parameter Estimation

Above, we have discussed many different approximations to the likelihood \mathcal{L} . Here we discuss finding the parameters that maximize this likelihood (minimize the $\chi^2 \equiv -2 \ln \mathcal{L}$). We then apply our methods to estimating the power in discrete bins of ℓ . This application provides another demonstration of the importance of using a better approximation to the likelihood than a Gaussian.

The likelihood functions above depend on \mathcal{C}_ℓ which may in turn depend on other parameters, a_p , which are, e.g., the physical parameters of a theory. If we write the parameters as $a_p + \delta a_p$ we can find the correction, δa_p , that minimizes χ^2 by solving

$$\delta a_p = -\frac{1}{2} \mathcal{F}_{pp'}^{-1} \frac{\partial \chi^2}{\partial a_{p'}}, \quad (48)$$

where

$$\mathcal{F}_{pp'} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_p \partial a_{p'}} \quad (49)$$

is the curvature matrix for the parameters a_p . If the χ^2 were quadratic (i.e., Gaussian \mathcal{L}) then Eq. 48 would be exact. Otherwise, in most cases, near enough to its minimum, χ^2 is approximately quadratic and an iterative application of Eq. 48 converges quite rapidly. The covariance matrix for the uncertainty in the parameters is given by $\langle \delta a_p \delta a_{p'} \rangle = \mathcal{F}_{pp'}^{-1}$. This is just an approximation to the Newton-Raphson technique for finding the root of $\partial \mathcal{L} / \partial a_p = 0$; a similar technique is used in quadratic estimation of \mathcal{C}_ℓ [13,7,9].

As our worked example here, we parameterize the power spectrum by the power in $B = 1$ to 11 bins, \mathcal{C}_B . Within each of the bins, we assume $\mathcal{C}_\ell = \mathcal{C}_B$ to be independent of ℓ . We have chosen the offset lognormal approximation. The χ^2 completely describes the model:

$$\chi^2 = \sum_{i,j} (Z_i^t - Z_i^d) M_{ij}^Z (Z_j^t - Z_j^d) + \chi_{\text{cal}}^2; \quad (50)$$

$$\chi_{\text{cal}}^2 \equiv \sum_\alpha \frac{(u_\alpha - 1)^2}{\sigma_\alpha^2}; \quad (51)$$

$$Z_i^d \equiv \ln(D_i + x_i); \quad (52)$$

$$Z_i^t \equiv \ln \left(\sum_B u_{\alpha(i)} f_{iB} \mathcal{C}_B + x_i \right); \quad (53)$$

$$M_{ij}^Z \equiv M_{ij} (D_i + x_i) (D_j + x_j) \quad \text{no sum}; \quad (54)$$

where M_{ij} is the weight matrix for the band powers D_i . We have modeled the signal contribution to the data, D_i , as an average over the power spectrum, $\sum_B f_{iB} \mathcal{C}_B$, times a calibration parameter, $u_{\alpha(i)}$. For simplicity, we take the prior probability distribution for this parameter to be normally distributed. Since the datasets have already been calibrated, the mean of this distribution is at $u_\alpha = 1$. The calibration parameter index, α is a function of i since different power spectrum constraints from the same dataset all share the same calibration uncertainty. We solve simultaneously for the u_α and \mathcal{C}_B ; i.e., together they form the set of parameters, a_p , in Eq. 48. For those experiments reported as band-powers together with the trace of the window function, W_ℓ^i , the filter is taken to be

$$f_{iB} = \frac{\sum_{\ell \in B} W_\ell^i / \ell}{\sum_\ell W_\ell^i / \ell}. \quad (55)$$

For Saskatoon and COBE/DMR, our D_i are themselves estimates of the power in bands. For these cases the above equation applies, but with W_ℓ / ℓ set to a constant within the estimated band and zero outside. The estimated bands have different ℓ ranges than the target bands.

Instead of the curvature matrix of Eq. 49 we use an approximation to it that ignores a logarithmic term. Including this term can cause the curvature matrix to be non-positive definite as the iteration proceeds. The approximation has no influence on our determination of the best fit power spectrum, but does affect the error bars. We expect that the effect is quite small.

We now proceed to find the best-fit power spectrum given different assumptions about the value of x_i , binnings of the power spectrum and editings of the data.

We have determined the x_i only for COBE/DMR, Saskatoon, SP89, OVRO7 and SuZIE. To test the sensitivity to the unknown x_i s we found the minimum- χ^2 power spectrum assuming the two extremes of $x_i = 0$ (corresponding to lognormal) and $x_i = \infty$ (corresponding to Gaussian). These two power spectra are shown in Fig. 4. Note that both power spectra were derived using our measured x_i values; only the unknown x_i values were varied. The variation in the results would be much greater if we let these x_i values be at their extremes.

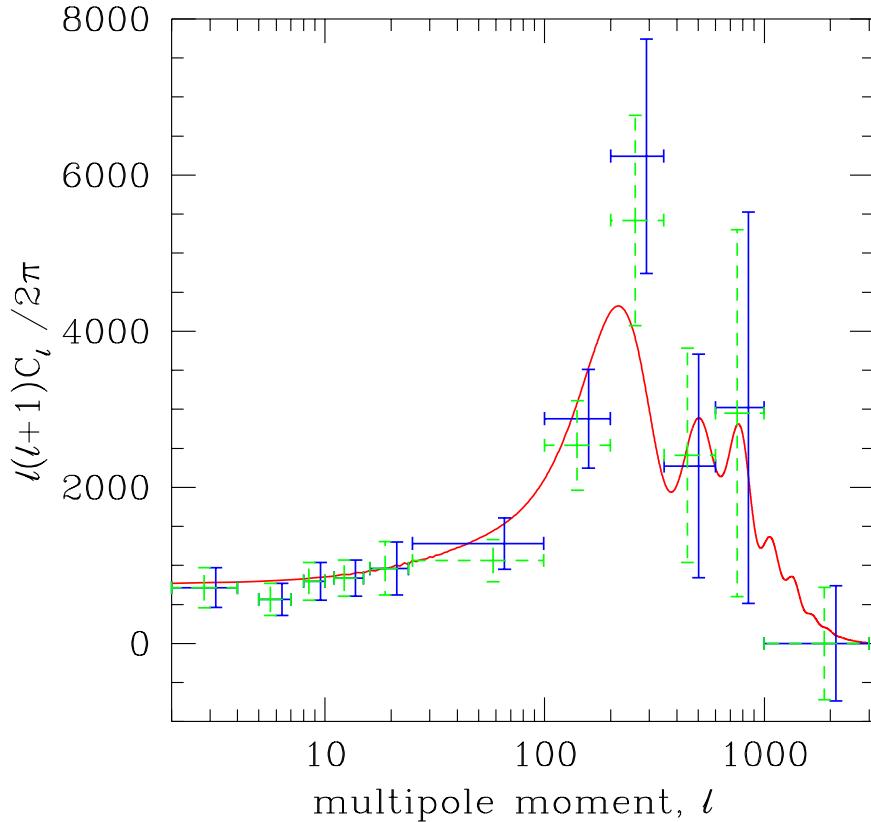


FIGURE 4. Power spectra that minimize the χ^2 in Eq. 50. The solid (dashed) error bars assume $x = 0$ ($x = \infty$) for those datasets with no determination of x ; the two sets have been offset slightly for display purposes. Solid curve is standard CDM.

V CONCLUSIONS

In this proceedings, we have discussed the beginning and end of the CMB data-analysis process, and shown how they can be unified in a Bayesian likelihood formalism. We have introduced new techniques and approximations that

- allow simultaneous calculation of the noise properties and underlying sky map.
- approximate the full cosmological likelihood function with minimal information.

Crucially, both of these innovations can be calculated in relatively small or comparable amounts of time to the most computationally-challenging aspects of the problem. Hence, they await the same innovations necessary to perform these calculations in the first place for upcoming megapixel datasets [6].

ACKNOWLEDGEMENTS

AHJ would like to thank Mike Seifert, Julian Borrill, the COMBAT collaboration, and the Saskatoon, MAXIMA and BOOMERANG teams for many helpful conversations during the course of this research. AHJ acknowledges support by NASA grants NAG5-6552 and NAG5-3941 and by NSF cooperative agreement AST-9120005.

REFERENCES

1. Wright, E.L, astro-ph/9612105.
2. Tegmark, M. *Astroph. Jour.* **455**, 1 (1996); Tegmark, M. *Phys. Rev.* **D56**, 4514 (1997).
3. Ferreira, P.G. & Jaffe, A.H., in preparation (1999).
4. See, e.g., Eisenstein, D.J., Hu, W. & Tegmark, M. 1998 astro-ph/9807130; Efstathiou & Bond, 1998, astro-ph/9807103.
5. Bond, J.R., Jaffe, A.H. & Knox, L.E., *Astrophys. J.*, submitted (1999).
6. Bond, J.R., Crittenden, R.C., Jaffe, A.H. & Knox, L.E., *Computers in Science & Engineering*, invited review (1999) and references therein.
7. Bond, J.R., Jaffe, A.H. & Knox, L.E., *Phys. Rev.* **D57**, 2117 (1998) and references therein.
8. Seljak, U. 1997, astro-ph/9710269.
9. Oh, S.P., Spergel, D.N., and Hinshaw, G. 1998, astro-ph/9805339.
10. Netterfield, C.B., Devlin, M.J., Jarosik, N., Page, L., and Wollack, E.J. 1997, *Astrophys. J.*, 474, 47.
11. Wollack, E.J., Devlin, M.J., Jarosik, N., Netterfield, C.B., Page, L., and Wilkinson, D. 1997, *Astrophys. J.*, 476, 440.
12. Leitch, E. 1998, Caltech PhD. thesis.
13. Tegmark, M. 1997, *Phys. Rev.* **D55**, 5895; astro-ph/9611174.

